

Some Case Studies in the use of Statistical tools

Mrs. Sahana Prasad

Associate Professor, Department of Statistics, Christ University, Bangalore, India

Abstract: The world is producing more statistical information than ever before and many people feel overwhelmed by it. It is important to know what information we should pay attention to, and what to do with these and how to use them. Since political, economic and other social conditions are rapidly changing, a researcher has to be equipped well to deal with them and use the appropriate tools with caution. This requires that the researcher has enough exposure and familiarity with different tools and techniques as the onus of selecting and implementing the right one rests with the researcher. In order to fulfil this responsibility, the researcher has to pay attention to every small details of the study, right from data collection, measurement of data, using the right tool for different types of data and making the correct inferences from the data using the correct and valid techniques of analysis. A very important aspect in this is that the researcher should become conversant with the available tools and the situations in which they can be used.

Keywords: Statistical tools, data, suitability of tools, case studies

I. INTRODUCTION

In earlier days, use of statistics was generally confined to techniques of organizing and classifying data which were collected by observation in a survey or an experiment. More emphasis was on finding the essential characteristics of data. One of the major hurdles which researchers faced was the difficulty in using numbers and the time required in manual computation which has been of course overcome by the introduction of electronic calculators and software packages. A well planned program is essential before collecting the data. Knowledge about alternative statistical methods, their assumptions, limitations and their usability is very important to the researcher. Enough exposure and familiarity with different tools and techniques is very important for selecting and implementing the right tool.

Case Studies showing the use of Statistical tools

Case Study 1: A common measure of the relative value of a company's stock is the price to earnings ratio...A relatively low PE indicates either a relatively under-valued stock or a company expected to have low or even negative future earnings growth, while a relatively high PE indicates either an over-valued stock or a company expected to have robust future earnings growth.

Consider the following companies, for which we want to determine the aggregate PE:

Company	PE Ratio
A	22.50
B	24.20
C	20.00

Right tool: Harmonic Mean

Reason: We are averaging ratios and not numbers.

Note: If the companies have similar market values, then the average PE would be the harmonic mean. But, if the companies' market values are not similar, we need to use weighted HM. Suppose that the companies' market values were as follows:

Company	Market Capitalization	PE Ratio
A	5,000,000	22.50
B	2,000,000	24.20
C	200,000	20.00
D	20,000	60.00

Observe that Company A comprises a big chunk of the total market value of the four companies, and as so, its PE should be given greater weight. Thus, weighted harmonic mean will be the right tool.

Case Study 2: Health Care: Diabetes People with diabetes may develop other health complications associated with the disease. The following information is based on a feature in USA Today entitled "A Look at Statistics That Shape Our Lives." About 40% of all people with diabetes will also develop hypertension (blood pressure problems), and about 30% of people with diabetes will develop an eye disease. Suppose that you are the director of a health care center that has 0 people with diabetes and no other related.

health problems. Part of your job is to monitor these patients for symptoms of new illness related to diabetes so that corrective measures can be started. What is the probability that

- None of the diabetes patients will ever develop related hypertension?*
- Fewer than 5 of the diabetes patients will ever develop related hypertension?*
- No more than 2 of the diabetes patients will ever develop a related eye disease?*
- At least 6 of the diabetes patients will never develop a related eye disease?*

Suitable Tool: Binomial distribution

Reason: The researcher can now appreciate why binomial distribution is used

Case Study 3: A random sample of college students were asked if they supported the 'Free Economy' move initiated by Dr Manmohan Singh in 1991. The same group of students was re-interviewed in 2000 to see if the economic liberalization policy followed so far helped change their earlier stand on the issue. The data are cast in a 2x2 table. Have the changes in the intervening period helped change their earlier stand?

Suitable Tool: McNemar test

Reason: The test is appropriate because the samples are related and the data are in nominal scale.

Case Study 4 : A researcher randomly selects 80 parents to get their opinion on introduction of "Sex education" classes in high school. After taking their opinion, they were given material highlighting the importance of such classes to children. A survey was conducted again and the results were tabulated in the form of table given below:

Opinion before reading material	Opinion after reading material	
	Favour of classes	Not in favour of classes
	Favour of classes	30
Not in favour of classes	34	02

Can the researcher draw conclusions whether there was a significant change in the attitude of parents after reading the given material?

Suitable tool: This requires McNemar's test as

- (i) The subjects are drawn randomly drawn
- (ii) The collected data can be grouped into "before- after" categories
- (iii) Data is nominal
- (iv) No assumptions are made about normality, continuity or equality/ in-equality of variances

Case Study 5:

In an industrial production line, items are inspected periodically for defectives. The following is a sequence of defective items, D, and non-defective items, N, produced by the production line: D D N N N D N N D D N N N N N D D D N N D N N N N D N D. Use the large sample theory for the runs-test, with a significance level of 0.05, to determine whether the defectives are occurring at random.

Suitable Tool: One sample runs test for large sample

Reason: The null hypothesis refers to the randomness of a single group of events.

Case Study 6:

Three cricket teams completed four 20-over matches. The following table shows how the teams fared in the 4 innings they played. Do the three groups differ in their average performances? Teams

A	B	C
150(7)	140(5)	170(11)
122(2)	120(1)	170(11)
130(3)	132(4)	145(6)
160(9)	170(11)	145(8)

Suitable Tool: Kruskal-Wallis one-way ANOVA

Reason: The data requires a test for tackling simultaneously 3 independent samples. Cricket scores can easily be expressed in ordinal measurements. Kruskal-Wallis one-way ANOVA seems appropriate for testing the differences.

Case Study 7: McDonald and Kreitman (1991) sequenced the alcohol dehydrogenase gene in several individuals of three species of Drosophila. Varying sites were classified as synonymous (the nucleotide variation does not change an amino acid) or amino acid replacements, and they were also classified as polymorphic (varying within a species) or fixed differences between species. The two nominal variables are thus synonymicity ("synonymous" or "replacement") and fixity ("polymorphic" or "fixed"). In the absence of natural selection, the ratio of synonymous to replacement sites should be the same for polymorphisms and fixed differences. There were 43 synonymous polymorphisms, 2 replacement polymorphisms, 17 synonymous fixed differences, and 7 replacement fixed differences.

	synonymous	replacement
polymorphisms	43	2
fixed	17	7

Suitable tool: Fisher's exact test

Wrong tool: Chi square test

Reason: Since the expected numbers in some classes are small, the chi-squared test will give inaccurate results.

ACKNOWLEDGMENT

I wish to thank Christ University, Bangalore, India for funding a research project on use of Statistical tools. This paper has been drawn from the monograph which is an outcome of this research.

REFERENCES

- [1] B Y Brase, C. H. (n.d.). *Understanding basic statistics*.
- [2] Healey, J. (n.d.). *Statistics- A tool for Social Research*. Newport University.
- [3] Howell, C. D. (n.d.). *Statistical Methods for Psychology*.
- [4] Linneman, T. J. (n.d.). *Social Statistics- Basics and Beyond*.
- [5] Majumdar, P. (n.d.). *Applied Statistics- A course for social sciences*. Rawat Publishers.
- [6] P.K.Majumdar. (n.d.). *Statistics- A tool for Social Sciences*.